

## STIC Biotechnology Systems Branch

### RAW SEQUENCE LISTING ERROR REPORT

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following computer readable form:

Application Serial Number: 10/506,630  
Source: PT/10  
Date Processed by STIC: 4/21/05

THE ATTACHED PRINTOUT EXPLAINS DETECTED ERRORS.

PLEASE FORWARD THIS INFORMATION TO THE APPLICANT BY EITHER:

- 1) INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE APPLICANT, WITH A NOTICE TO COMPLY or,
- 2) TELEPHONING APPLICANT AND FAXING A COPY OF THIS PRINTOUT, WITH A NOTICE TO COMPLY

FOR CRF SUBMISSION AND PATENTIN SOFTWARE QUESTIONS, PLEASE CONTACT MARK SPENCER, TELEPHONE: 571-272-2510; FAX: 571-273-0221

TO REDUCE ERRORED SEQUENCE LISTINGS, PLEASE USE THE CHECKER VERSION 4.2.2 PROGRAM, ACCESSIBLE THROUGH THE U.S. PATENT AND TRADEMARK OFFICE WEBSITE. SEE BELOW FOR ADDRESS:

<http://www.uspto.gov/web/offices/pac/checker/chkrnote.htm>

Applicants submitting genetic sequence information electronically on diskette or CD-Rom should be aware that there is a possibility that the disk/CD-Rom may have been affected by treatment given to all incoming mail.

Please consider using alternate methods of submission for the disk/CD-Rom or replacement disk/CD-Rom.

Any reply including a sequence listing in electronic form should NOT be sent to the 20231 zip code address for the United States Patent and Trademark Office, and instead should be sent via the following to the indicated addresses:

1. EFS-Bio (<<http://www.uspto.gov/ebs/efs/downloads/documents.htm>> , EFS Submission User Manual - ePAVE)
2. U.S. Postal Service: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450
3. Hand Carry, Federal Express, United Parcel Service, or other delivery service (EFFECTIVE 01/14/05):  
U.S. Patent and Trademark Office, Mail Stop Sequence, Customer Window, Randolph Building, 401 Dulany Street, Alexandria, VA 22314

Revised 01/24/05

BEST AVAILABLE COPY

## Raw Sequence Listing Error Summary

### ERROR DETECTED

### SUGGESTED CORRECTION

SERIAL NUMBER: 10/506,630

ATTN: NEW RULES CASES: PLEASE DISREGARD ENGLISH "ALPHA" HEADERS, WHICH WERE INSERTED BY PTO SOFTWARE

- 1 ☐ **Wrapped Nucleics  
Wrapped Aminos** The number/text at the end of each line "wrapped" down to the next line. This may occur if your file was retrieved in a word processor after creating it. Please adjust your right margin to .3; this will prevent "wrapping."
- 2 ☐ **Invalid Line Length** The rules require that a line not exceed 72 characters in length. This includes white spaces.
- 3 ☒ **Misaligned Amino  
Numbering** The numbering under each 5<sup>th</sup> amino acid is misaligned. Do not use tab codes between numbers; use space characters, instead.
- 4 ☐ **Non-ASCII** The submitted file was not saved in ASCII(DOS) text, as required by the Sequence Rules. Please ensure your subsequent submission is saved in ASCII text.
- 5 ☐ **Variable Length** Sequence(s) contain n's or Xaa's representing more than one residue. Per Sequence Rules, each n or Xaa can only represent a single residue. Please present the maximum number of each residue having variable length and indicate in the <220>-<223> section that some may be missing.
- 6 ☐ **PatentIn 2.0  
"bug"** A "bug" in PatentIn version 2.0 has caused the <220>-<223> section to be missing from amino acid sequences(s). Normally, PatentIn would automatically generate this section from the previously coded nucleic acid sequence. Please manually copy the relevant <220>-<223> section to the subsequent amino acid sequence. This applies to the mandatory <220>-<223> sections for Artificial or Unknown sequences.
- 7 ☐ **Skipped Sequences  
(OLD RULES)** Sequence(s) missing. If intentional, please insert the following lines for each skipped sequence:  
(2) INFORMATION FOR SEQ ID NO:X: (insert SEQ ID NO where "X" is shown)  
(i) SEQUENCE CHARACTERISTICS: (Do not insert any subheadings under this heading)  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:X: (insert SEQ ID NO where "X" is shown)  
This sequence is intentionally skipped  
  
Please also adjust the "(ii) NUMBER OF SEQUENCES:" response to include the skipped sequences.
- 8 ☐ **Skipped Sequences  
(NEW RULES)** Sequence(s) missing. If intentional, please insert the following lines for each skipped sequence.  
<210> sequence id number  
<400> sequence id number  
000
- 9 ☐ **Use of n's or Xaa's  
(NEW RULES)** Use of n's and/or Xaa's have been detected in the Sequence Listing.  
Per 1.823 of Sequence Rules, use of <220>-<223> is MANDATORY if n's or Xaa's are present.  
In <220> to <223> section, please explain location of n or Xaa, and which residue n or Xaa represents.
- 10 ☐ **Invalid <213>  
Response** Per 1.823 of Sequence Rules, the only valid <213> responses are: Unknown, Artificial Sequence, or scientific name (Genus/species). <220>-<223> section is required when <213> response is Unknown or is Artificial Sequence
- 11 ☐ **Use of <220>** Sequence(s) missing the <220> "Feature" and associated numeric identifiers and responses.  
Use of <220> to <223> is MANDATORY if <213> "Organism" response is "Artificial Sequence" or "Unknown." Please explain source of genetic material in <220> to <223> section.  
(See "Federal Register," 06/01/1998, Vol. 63, No. 104, pp. 29631-32) (Sec. 1.823 of Sequence Rules)
- 12 ☐ **PatentIn 2.0  
"bug"** Please do not use "Copy to Disk" function of PatentIn version 2.0. This causes a corrupted file, resulting in missing mandatory numeric identifiers and responses (as indicated on raw sequence listing). Instead, please use "File Manager" or any other manual means to copy file to floppy disk.
- 13 ☐ **Misuse of n/Xaa** "n" can only represent a single nucleotide; "Xaa" can only represent a single amino acid

AMC - Biotechnology Systems Branch - 09/09/2003

BEST AVAILABLE COPY



PCT

## RAW SEQUENCE LISTING

DATE: 04/21/2005

PATENT APPLICATION: US/10/506,630

TIME: 09:58:27

Input Set : A:\30062-20046.01 - Seqlist (from parent .45).txt

Output Set: N:\CRF4\04212005\J506630.raw

4 <110> APPLICANT: GOKHALE, Rajesh  
 5 TSUJI, Stuart  
 6 KHOSLA, Chaitan  
 7 WU, Nicholas  
 8 CANE, David  
 10 <120> TITLE OF INVENTION: METHODS TO MEDIATE POLYKETIDE SYNTHASE  
 11 MODULE EFFECTIVENESS  
 13 <130> FILE REFERENCE: 300622004601  
 15 <140> CURRENT APPLICATION NUMBER: US 10/506,630  
 C--> 16 <141> CURRENT FILING DATE: 2004-09-03  
 18 <150> PRIOR APPLICATION NUMBER: PCT/US03/06910  
 19 <151> PRIOR FILING DATE: 2002-03-04  
 21 <150> PRIOR APPLICATION NUMBER: US 10/091,244  
 22 <151> PRIOR FILING DATE: 2002-03-04  
 24 <150> PRIOR APPLICATION NUMBER: 60/361,758  
 25 <151> PRIOR FILING DATE: 2002-03-04  
 27 <160> NUMBER OF SEQ ID NOS: 41  
 29 <170> SOFTWARE: FastSEQ for Windows Version 4.0

Does Not Comply  
Corrected Diskette Needed

pp 1,4

## ERRORED SEQUENCES

55 <210> SEQ ID NO: 3  
 56 <211> LENGTH: 31  
 57 <212> TYPE: PRT  
 58 <213> ORGANISM: Artificial Sequence  
 60 <220> FEATURE:  
 61 <223> OTHER INFORMATION: N-terminal linker of M3  
 63 <400> SEQUENCE: 3  
 64 Met Thr Asp Ser Glu Lys Val Ala Glu Tyr Leu Arg Arg Ala Thr Leu  
 E--> 65 1 5 10 15  
 66 Asp Leu Arg Ala Ala Arg Gln Arg Ile Arg Glu Leu Glu Ser Asp  
 E--> 67 20 25 30  
 238 <210> SEQ ID NO: 18  
 239 <211> LENGTH: 18  
 240 <212> TYPE: PRT  
 241 <213> ORGANISM: Artificial Sequence  
 243 <220> FEATURE:  
 244 <223> OTHER INFORMATION: Intra-polypeptide linker  
 246 <400> SEQUENCE: 18  
 247 Gly Gly Ala Thr Gly Ala Glu Gln Ala Ala Pro Ala Thr Thr Ala Pro  
 E--> 248 1 5 10 15  
 249 Val Asp

give source of genetic material  
(see item 11 on

Error Summary

Sheet)

misaligned amino acid  
numbers  
(see item 3 on Error Summary  
Sheet)

misaligned nos.

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/506,630

DATE: 04/21/2005

TIME: 09:58:27

Input Set : A:\30062-20046.01 - Seqlist (from parent .45).txt

Output Set: N:\CRF4\04212005\J506630.raw

1369 <210> SEQ ID NO: 41  
 1370 <211> LENGTH: 1099  
 1371 <212> TYPE: PRT  
 1372 <213> ORGANISM: Artificial Sequence  
 1374 <220> FEATURE:  
 1375 <223> OTHER INFORMATION: Alignment of the EryA SU  
 1377 <400> SEQUENCE: 41

R4

1378	Lys	Asp	Ala	Asp	Asp	Ile	Ile	Gly	Met	Ala	Phe	Gly	Val	His	Asn	Gly
1379	1				5					10					15	
1380	Glu	Leu	Glu	Phe	Ile	Val	Gly	Arg	Gly	Asp	Ala	Val	Thr	Glu	Met	Thr
1381				20					25					30		
1382	Asp	Leu	Asp	Ala	Leu	Phe	Asp	Pro	Asp	Pro	Gln	Arg	His	Gly	Thr	Ser
1383			35					40					45			
1384	Tyr	Ser	Arg	His	Ala	Phe	Leu	Asp	Gly	Ala	Asp	Ala	Ala	Ile	Ser	Leu
1385		50					55					60				
1386	Met	Gln	Val	Thr	Thr	Leu	Phe	Asn	Ile	Asp	His	Ser	Arg	Gly	Ser	Asp
1387	65					70				75					80	
1388	Leu	Ala	Ala	Tyr	Gln	Gly	Gly	Gln	Asp	Ala	Val	Val	Pro	Glu	Asp	Ser
1389					85					90					95	
1390	Glu	Leu	Leu	Thr	Asn	Ser	Ser	Ala	Val	Val	Ala	Val	Leu	Leu	Ala	Val
1391				100					105					110		
1392	Thr	Ser	Val	Ala	Leu	Ser	Cys	Gly	Ser	Asp	Gly	Asp	Cys	Gly	Leu	Val
1393			115					120					125			
1394	Ala	Val	Ser	Ala	Gly	Glu	Val	Phe	Thr	Glu	Ser	Arg	Gln	Gly	Gly	Ala
1395		130					135					140				
1396	Val	Cys	Ala	Ser	Ala	Glu	Asp	Gly	Phe	Ala	Val	Ala	Val	Val	Leu	Gln
1397	145					150				155					160	
1398	Arg	Asp	Arg	Ala	Arg	Gln	Gly	Val	Val	Ala	Ser	Leu	Gln	Ala	Ser	Val
1399				165					170						175	
1400	Ala	Gln	Arg	Lys	Trp	Ala	Arg	Ala	Ile	Thr	Gly	Ala	Val	Ala	Val	Arg
1401			180						185				190			
1402	Val	Ala	Ser	Leu	Ala	Thr	Lys	Ser	Gly	Ser	Ser	Gly	Val	Leu	Leu	Ser
1403			195					200					205			
1404	Ile	Ala	Val	Ala	Ile	Val	Val	Gly	Leu	Asn	Arg	Leu	Val	Pro	Met	Cys
1405		210					215					220				
1406	Arg	Gly	Arg	Ser	Pro	Leu	Ile	Glu	Ser	Ser	Gly	Val	Glu	Leu	Ala	Glu
1407	225					230					235				240	
1408	Ala	Val	Ser	Pro	Pro	Pro	Ala	Ala	Asp	Gly	Val	Gly	Ala	Val	Val	Ile
1409				245						250					255	
1410	Ala	Pro	Glu	Pro	Glu	Pro	Leu	Pro	Glu	Pro	Gly	Pro	Val	Gly	Val	Leu
1411				260					265					270		
1412	Ala	Ala	Ala	Asn	Ser	Val	Val	Leu	Leu	Ala	Arg	Thr	Glu	Thr	Ala	Leu
1413			275					280					285			
1414	Ala	Arg	Leu	Leu	Glu	Ser	Ala	Val	Asp	Asp	Ser	Val	Pro	Leu	Thr	Ala
1415		290					295					300				
1416	Leu	Ala	Ser	Ala	Leu	Thr	Gly	Ala	His	Leu	Pro	Arg	Ala	Leu	Ile	Ala
1417	305					310					315				320	
1418	Gly	Asp	His	Glu	Gln	Leu	Arg	Gly	Gln	Arg	Ala	Val	Glu	Val	Ala	Ala
1419				325					330						335	

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/506,630

DATE: 04/21/2005

TIME: 09:58:27

Input Set : A:\30062-20046.01 - Seqlist (from parent .45).txt

Output Set: N:\CRF4\04212005\J506630.raw

```

1420 Pro Gly Ala Thr Thr Gly Thr Ala Ser Ala Gly Gly Val Val Phe Val
1421           340           345           350
1422 Ala Glu Ala Arg Gly Ser Val Pro Glu Ser Ile Ala Glu Asp Ala Val
1423           355           360           365
1424 Leu Ser Glu Val Ala Gly Phe Ser Ala Ser Glu Val Leu Glu Gln Arg
1425           370           375           380
1426 Pro Asp Ala Pro Ser Leu Glu Asp Val Val Leu Ser Val Met Val Arg
1427 385           390           395           400
1428 Leu Gly Ala Cys Val Ser Ser Ile Gln Ile Val Ala Val Leu Ser Leu
1429           405           410           415
1430 Glu Asp Gly Val Arg Val Val Ala Leu Arg Ala Lys Ala Leu Arg Ala
1431           420           425           430
1432 Leu Ala Gly Arg Gly Gly Val Ser Leu Ala Ala Pro Gly Glu Arg Ala
1433           435           440           445
1434 Arg Ala Leu Ile Ala Pro Trp Glu Asp Arg Ile Ser Val Ala Ser Ser
1435           450           455           460
1436 Ser Val Val Val Ser Asp Pro Glu Ala Leu Ala Glu Leu Val Ala Arg
1437 465           470           475           480
1438 Cys Glu Asp Glu Gly Val Arg Ala Lys Thr Leu Pro Asp Tyr Ser Ser
1439           485           490           495
1440 Arg His Val Glu Glu Ile Arg Glu Thr Ile Leu Ala Asp Leu Asp Gly
1441           500           505           510
1442 Ile Ser Ala Arg Arg Ala Ala Ile Pro Leu Tyr Thr Leu His Gly Glu
1443           515           520           525
1444 Arg Arg Asp Gly Ala Asp Met Gly Pro Arg Tyr Asp Asn Leu Ser Gln
1445           530           535           540
1446 Arg Asp Glu Val Ser Ala Ala Val Ala Asp His Ala Thr Val Met Pro
1447 545           550           555           560
1448 Val Ile Thr Ala Ala Val Gln Glu Ile Ala Ala Asp Ala Val Ala Ile
1449           565           570           575
1450 Gly Ser His Asp Thr Ala Glu Glu His Ile Ile Ala Glu Leu Ala Arg
1451           580           585           590
1452 His Val His Ala Val Asp Trp Arg Asx Val Phe Pro Ala Ala Pro Pro
1453           595           600           605
1454 Val Ala Asn Glu Pro Gln Tyr Leu Ala Pro Glu Val Ser Gln Leu Ala
1455           610           615           620
1456 Asp Ser Arg Arg Val Asp Arg Pro Leu Ala Thr Thr Pro Val Asp Leu
1457 625           630           635           640
1458 Glu Gly Gly Phe Val His Gly Ser Ala Pro Glu Ser Leu Thr Ser Ala
1459           645           650           655
1460 Val Glu Lys Ala Gly Arg Val Val Pro Val Ala Ser Ala Asp Arg Glu
1461           660           665           670
1462 Ala Ser Ala Ala Leu Arg Glu Val Pro Gly Glu Val Ala Gly Leu Val
1463           675           680           685
1464 His Thr Gly Ala Ala Thr His Leu Ala His Gln Ser Leu Gly Glu Ala
1465           690           695           700
1466 Gly Val Arg Ala Pro Leu Trp Leu Val Ser Arg Ala Leu Gly Glu Ser
1467 705           710           715           720
1468 Glu Pro Val Asp Pro Glu Gln Ala Met Val Trp Leu Gly Arg Val Met

```

## RAW SEQUENCE LISTING

DATE: 04/21/2005

PATENT APPLICATION: US/10/506,630

TIME: 09:58:27

Input Set : A:\30062-20046.01 - Seqlist (from parent .45).txt

Output Set: N:\CRF4\04212005\J506630.raw

```

1469          725          730          735
1470 Gly Leu Thr Pro Glu Arg Trp Gly Gly Leu Val Leu Pro Ala Glu Pro
1471          740          745          750
1472 Ala Pro Gly Asp Gly Glu Ala Phe Val Cys Leu Gly Ala Asp Gly His
1473          755          760          765
1474 Glu Asp Gln Val Ile Asp His Ala Arg Tyr Gly Arg Leu Val Arg Ala
1475          770          775          780
1476 Pro Leu Gly Thr Arg Glu Ser Ser Trp Glu Ala Ala Val Gly Thr Gly
1477 785          790          795          800
1478 Ala Leu Gly His Val Ala His Ala Arg Cys Val Glu Asp Val Val Ser
1479          805          810          815
1480 Arg Arg Gly Val Asp Ala Pro Gly Ala Ala Glu Glu Ala Glu Leu Val
1481          820          825          830
1482 Ala Ile Ala Lys Thr Thr Ile Thr Cys Val Ala Asp Arg Glu Gln Leu
1483          835          840          845
1484 Ser Lys Leu Leu Glu Glu Leu Arg Gly Gln Gly Arg Pro Val Arg Thr
1485          850          855          860
1486 Val Val Thr Ala Gly Val Pro Glu Ser Arg Pro Leu His Glu Ile Gly
1487 865          870          875          880
1488 Glu Leu Glu Ser Val Cys Ala Ala Val Thr Gly Ala Arg Leu Asp Leu
1489          885          890          895
1490 Cys Pro Asp Ala Glu Thr His Val Leu Phe Gly Gly Val Trp Gly Ser
1491          900          905          910
1492 Ala Asn Leu Gly Ala Ser Ala Ala Asn Ala Tyr Asp Ala Leu His Arg
1493          915          920          925
1494 Arg Ala Glu Arg Ala Ala Thr Ser Val Ala Gly Ala Gly Glu Gly Met
1495          930          935          940
1496 Ala Thr Gly Asp Leu Glu Gly Leu Thr Arg Arg Leu Arg Pro Met Ala
1497 945          950          955          960
1498 Pro Glu Arg Ile Arg Ala Leu His Gln Ala Asp Asn Gly Asp Thr Cys
1499          965          970          975
1500 Val Ser Ile Ala Asp Val Asp Trp Glu Ala Phe Ala Val Gly Phe Thr
1501          980          985          990
1502 Ala Arg Pro Arg Pro Leu Leu Asp Glu Leu Val Thr Pro Ala Val Gly
1503          995          1000          1005
1504 Ala Val Pro Ala Val Gln Ala Pro Ala Arg Glu Met Thr Ser Gln Glu
1505          1010          1015          1020
1506 Leu Glu Phe Thr His Ser His Val Ala Ile Leu His Ser Ser Pro Asp
1507 1025          1030          1035          1040
1508 Ala Val Gly Gln Asp Gln Pro Thr Glu Phe Leu Thr Val Gly Asn Gln
1509          1045          1050          1055
1510 Leu Gln Gln Ala Thr Leu Ala Leu Pro Ala Thr Leu Val Phe Glu Thr
1511          1060          1065          1070
1512 Val Arg Arg Leu Asp His Ile Gly Gln Gln Asp Ser Gly Thr Pro Ala
1513          1075          1080          1085
1514 Arg Glu Ala Ser Ser Ala Leu Arg Asp Gly Tyr
1515          1090          1095

```

E--&gt;

1549

(25) delete

## VERIFICATION SUMMARY

DATE: 04/21/2005

PATENT APPLICATION: US/10/506,630

TIME: 09:58:28

Input Set : A:\30062-20046.01 - Seqlist (from parent .45).txt

Output Set: N:\CRF4\04212005\J506630.raw

L:16 M:271 C: Current Filing Date differs, Replaced Current Filing Date

L:65 M:332 E: (32) Invalid/Missing Amino Acid Numbering, SEQ ID:3

M:332 Repeated in SeqNo=3

L:248 M:332 E: (32) Invalid/Missing Amino Acid Numbering, SEQ ID:18

L:1549 M:332 E: (32) Invalid/Missing Amino Acid Numbering, SEQ ID:41